DOCUMENT RESUME

ED 436 176 IR 019 801

AUTHOR	Lockee, Barbara B.; Burton, John K.; Cross, Lawrence H.
TITLE	No Comparison: Distance Education Finds a New Use for "No Significant Difference."
PUB DATE	1999-02-00
NOTE	8p.; In: Proceedings of Selected Research and Development Papers Presented at the National Convention of the Association for Educational Communications and Technology [AECT] (21st, Houston, TX, February 10-14, 1999); see IR 019 753.
PUB TYPE	Information Analyses (070) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Audiovisual Instruction; Comparative Analysis; *Distance Education; Educational History; *Educational Media; *Educational Technology; Evaluation Methods; Evaluation Problems; *Evaluation Research; Higher Education; Instructional Effectiveness; Intermode Differences; Literature Reviews; *Research Methodology; *Research Problems; Statistical Significance

ABSTRACT

This paper details the origins of the media comparison study methodology, its current use as an evaluation instrument in distance education, and recommendations for more stringent discrimination between research and evaluation in the field of distance learning. The following topics are discussed: (1) the history of research in instructional technology from 1968 to the present; (2) recent research in distance learning, including the difference between evaluation and research, and weaknesses in many instructional technology studies (e.g., failing to link the study to a robust theory, poor literature review, weak treatment implementation, measurement flaws, inadequate sample size, and poor analyses) that bias the research toward not finding a statistically significant difference; and (3) evaluation versus research in distance education, focusing on appropriate uses of media comparisons for distance program evaluation, as well as alternative methods and exemplary models. (Contains 55 references.) (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 176

NO COMPARISON: DISTANCE EDUCATION FINDS A NEW USE FOR "NO SIGNIFICANT DIFFERENCE"

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Zenor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Barbara B. Lockee
John K. Burton
Lawrence H. Cross
Virginia Tech

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

Recent exponential growth in the field of distance education has unfortunately not been matched with equal growth in a quality research base in terms of informing effective practice. In their 1996 analysis of the distance education literature, McIsaac & Gunawardena state that "much research has taken the form of program evaluation, descriptions of individual programs, brief case studies, institutional surveys and speculative reports" (1996, p. 421). In the past few years, a particular form of study regarding the effectiveness of distance learning has enjoyed increased publication. Instructional technology professionals will recognize the methodology behind these reports as the media comparison study, newly revived for use in a distance education setting. Media comparison studies have historically formed the basis of much research in distance education (McIsaac & Gunawardena, 1996; Schlosser & Anderson, 1994), but are lately becoming even more common. These studies are predictably plagued with the same design issues as their predecessors, however their "no difference" outcomes are being reported for politically different reasons. This paper details the origins of the media comparison study, its current use as an evaluation instrument in distance education, and recommendations for more stringent discrimination between research and evaluation in the field of distance learning.

History of research in instructional technology

Since the adoption of modern media for instructional purposes, innumerable attempts have been made to measure the effect that a given technology has on student achievement. Early in the history of electronic technologies like film and radio, educational researchers were driven to demonstrate that these revolutionary devices had a positive impact on learning (Saettler, 1968). The most common approach to attempt this investigation was the "media comparison" study, so named because of its strategy of comparing the learning outcomes of an experimental group receiving instructional content via one medium against the outcomes of a control group receiving the same content through a different medium. The most popular control group was the "traditional", or lecture format class, with the instructor serving as the delivery medium. Even though comparison studies were fairly simple in design, variations of the research strategy exist. Most maintained the "media" as the independent variable, but some used the same instructional method (e.g., presentation via live lecture vs. presentation via videotape of the same lecture) while others utilized different instructional methods and different media (presentation via lecture vs. problem solving via computer-based instruction) (Ross & Morrison, 1996).

The use of media comparison studies dates back to the origins of mediated instruction. Early researchers in audiovisual education worked diligently to control all aspects of such experiments so that results would maintain validity and comparisons would be fair. For example, McClusky & McClusky produced a "Comparison of six modes of presentation of subject matter". Two trial experiments were conducted comparing six methods of presentation for the content depicted in two separate lessons: film only, slides with subtitles only, photographic pictures with subtitles only, and each medium with a supplemental question and answer session. Experiments were carefully controlled so participants viewed slide and print images for exactly the same time as such images appeared in the film. Recall of content was measured for each group using multiple-choice tests. The outcomes reflect some of the earliest evidence of what Russell (1997) calls the non-significant difference (NSD) phenomenon: "These comparisons show such inconsistent results that the film, slide, and print appear to possess no distinct advantage one over the other as far as these particular experiments are concerned" (Russell, 1997,p. 257). And, as history repeats itself, the tendency was to try this research design with the advent of any newer technological innovation, with consistent production of the same non-significant results.

In 1983, Richard Clark, one of the most renowned critics of instructional technology research, detailed the problems inherent in media comparison studies and the improper assumptions about their outcomes. He stated that "these findings were incorrectly offered as evidence that different media were 'equally effective' as conventional means in promoting learning. No significant difference results simply suggest that changes in outcome scores did not result from any systematic differences in the treatments compared" (Clark, 1983, p. 447). Clark emphasized that media are merely the delivery mechanisms for instructional content and do not impact the learning process. This perspective spurred great debate within the field of instructional technology (Clark, 1994a; 1994b; Kozma, 1991; 1994a; 1994b; Jonassen, Campbell, & Davidson, 1994; Morrison, 1994; Reiser, 1994; Shrock, 1994; Tennyson,

IR019501

2

BEST COPY AVAILABLE

1994). While some contend that certain attributes of media can and do effect learning outcomes (Kozma, 1991; 1994a, 1994b), Clark maintains that it is instructional method that influences learning, not the delivery medium (Clark, 1983; 1994a; 1994b). While this debate will undoubtedly continue, the futility of comparison studies to measure the impact of media on learning is consistently recognized in the field of instructional technology (Ross & Morrison, 1996). A chronological collection of hundreds of such experiments can be found at http://www2.ncsu.edu/oit/nsdsplit.htm (Russell, 1997). The database serves as a reminder to researchers that comparative designs will continue to provide predictable, non-significant outcomes.

While the tendency is to compare learning outcomes via different media is to demonstrate the greater effectiveness of the newer medium, distance education comparison studies have given the argument a new twist. As evidenced in the following examples, the outcomes are now used to demonstrate that the distance-delivered instructional event is at least equal to the campus-based, face-to-face version. Kanner, Runyon, & Desiderato (1958) espoused this approach in less optimistic terms when summarizing their televised instruction research by stating that televised sessions were no more detrimental to classroom learning than face-to-face instruction.

Recent research in distance learning

As anyone involved in the support of distance education programming is aware, the resources required to deliver such programming to geographically and temporally dispersed learners are not inconsequential. Though cost-saving goals are often highlighted in plans for reaching new and different student markets, the front-end investments needed in course development, delivery infrastructures, teaching technologies, and support staff can be formidable (Keegan, 1996; Musial & Kampmueller, 1996). In analyzing the use of distance program evaluation data, Thorpe (1988) explains that administrators venturing into this new educational arena are expectedly anxious to use positive evaluation results to promote the desirable aspects of providing opportunities for remote student clientele. Increased access to such programming does not seem to serve as a satisfactory benefit for the implementation of distance education efforts. Stakeholders desire to demonstrate that participants in distance-delivered courses receive the same quality of instruction off-campus as those involved in the "traditional" classroom setting. What better way to determine the equality of experiences than to compare student achievement between the two groups? For example, according to Newlands & McLean (1996) "The calming of fears about the quality of distance learning has been assisted by evidence that, in terms of assessment, distance students perform as well as internal students..." (p. 289). One of the most prominent early works in teleconferencing training, Bridging the Distance (Monson, 1978), employs a collection of comparison studies for this very reason—to ensure soon-to-be distance educators that off-campus students will be just as academically successful as their campus-bound counterparts.

The guaranteed validation of equality of learner achievement has led to use of comparison studies about distance education in almost every imaginable discipline. The research design remains exactly the same as previously compared mediated experiences, the on-campus students serve as the control group, since their experience is unmediated, while the distant students provide the treatment group. For example, "The 38 South Carolina campus students, considered the control group in this report, completed either all or the majority of their degree programs in traditional classroom settings on the Columbia campus" (Douglas, 1996, p. 878). Repeatedly, outcomes are embodied in statements like "there was no real difference between grades of in-class and ITV students" (Fox, 1996, p. 362). Some authors offer additional analysis such as, "There were no differences between pre-and post-tests (which measure increases in knowledge) across sites. This demonstrates that the program was effective in increasing knowledge..." (Reiss, Cameon, Matthews, & Shenkman, 1996, p. 350). Behind conclusions such as these is the conviction that distance learners are engaged in an equally rigorous instructional experience even though they are not participating in campus-based education. General unawareness of the history of instructional technology research, especially the out-dated notion of the effectiveness of comparison studies, is exemplified in a recent sociology journal through the following excerpt:

> Of greatest concern to us is the absence of well-crafted comparison studies, that examine not simply student attitudes towards distance learning, but the actual knowledge and skills that students acquire from televised teaching. Ideally, this demonstration could involve the same instructor teaching two sections of the same course during the same school term, one exclusively by live instruction and the other only by distance learning (Thyer, Polk, & Gaudin, 1997, p. 367).

Instructional technology research journals are not exempt from the publication of such studies. In a 1996 issue of *Educational Technology Research & Development*, Whetzel, Felker, and Williams began their article "A Real World Comparison of the Effectiveness of Satellite Training and Classroom Training" with an analysis of research regarding the effectiveness of televised instruction, summing up the mixed results by re-stating Clark's (1994) view that "...any necessary teaching method can be delivered by many media with similar learning results (Whetzel, Felker, & Williams, 1996, p. 6). However, their study used a research design that compared the achievement of the on-site versus distant students:

> For the two courses in which satellite and classroom training were compared, an analysis of covariance (ANCOVA) was used to compare delivery modes for nonequivalent groups (satellite and classroom participants), using pretest score as the covariate and posttest score as the dependent variable (Whetzel et al., 1996, p. 10).

The use of media comparison studies in distance learning is not limited to higher education settings. Barry & Runyan (1995) assembled a "Review of Distance Education Studies in the U.S. Military" in which they cited eight "empirical studies that compared student achievement in distance learning courses to achievement in comparable resident courses" (p. 43). Their closing statement embraced the reliable non-significant difference results as proof that the U.S. military could safely continue to invest in the expansion of distance learning initiatives. Due to the expanded publication of studies like these, a distinction must be made between valid research in distance education and evaluation efforts for distance program confirmation.

Methodology analysis

Although we often use the terms interchangeably, as we have noted earlier, evaluation and research are not the same although they may share many methods.

Evaluation is practical and concerned with how to improve a product or whether to buy and use a product. Studies that compare one program or media against another are primarily evaluation. Evaluations seeks to find the programs that "work" more cheaply, efficiently, quickly, effectively, etc.

Research, on the other hand tends to be more concerned with testing theoretical concepts and constructs or with attempting to isolate variables to observe their contributions to a process or outcome (Moore, Myers, & Burton, 1994, p. 35)

Research studies generate hypotheses from theory. In the case of so called "hard" sciences such as physics, these predictions are usually a quantitative point value, magnitude, or form function which become point predictions (Meehl, 1967). Point predictions become easier to refute as measurement improves, that is, the better the measurement the more the hypothesis is exposed to rejection. (Indeed many replications involve changes in measures rather than "study" conditions). Rejection, in the case of a point prediction, is a *modus tollens* refutation (i.e. T -> E, not E, therefore not T) (Popper, 1968). Research in the "soft" sciences however, does not, indeed cannot, test point predictions. Rather, the logical compliment of the predicted outcome, the point-*null* hypothesis (there is *no* difference between, for example, two groups of participant's mean scores) is tested. Interestingly, the effect of this manipulation is that the theoretically derived hypothesis is not subjected to true *modus tollens* – it cannot be *logically* refuted (Meehl, 1967). The null hypothesis can be accepted but not embraced as true, yet it's acceptance does not refute the core hypothesis (Orey, Garrison, & Burton, 1989). For this reason (among others such as the "loose" connection between theory and variables), it is often the case that "soft" science researchers commonly speculate about what could have occurred when the null is accepted. Such *post hoc* speculation is permissible because the core hypothesis cannot be truly refuted. Such theories never actually die, researchers just sort of "…lose interest in the thing and pursue other endeavors" (Meehl, 1978, p. 807). Theory-based research studies are not good candidates to be "repurposed" as "negative" evidence that something didn't happen.

A "second level" problem relates to what Reeves (1993, 1995) among others has referred to as pseudoscience. Reeves offers nine characteristics of pseudoscientific studies and estimates that perhaps 60% - 70% of "empirical-quantitative" studies in the major instructional technology research journals suffer from two or more of these flaws. The bulk of these weaknesses such as failing to link the study to a robust theory, poor literature review, weak treatment implementation, measurement flaws, inadequate sample size, and poor analyses (Reeves, 1993; 1995)) bias the research towards *not* finding a statistically significant difference (Burton & Magliaro, 1988). In other words, bad science and bad designs can produce no differences.

The last two comments relate to both theoretical and evaluation comparative studies. The first comment is that when we test a hypothesis, we test not just the variable of interest, we also test the assumption of *ceteris paribus* (all things are assumed to be equal except for those conditions that are actually manipulated) (Orey et al., 1989). It is in a fact a "folk" version of *ceteris paribus* which researchers often resort to when they explain the failure to find a predicted statistically significant difference by resorting to differences in the sample or task that were outside of what was being manipulated. To the extent that *ceteris paribus* is not true, the results of the study (in either "direction") are suspect. With test such as ANOVA and ANCOVA, this relates to the assumption of homogeneity of variance. This assumption is often not tested because the tests are assumed to be robust to such violations (Thompson, 1993). Unfortunately, it does not appear to be true with ANCOVA (see, e.g. Keppel & Zedeck, 1989) and may not be true of ANOVA either (e.g. Wilcox, Charlin, & Thompson, 1986).

Second, in research or evaluation, measurement is a problem. Faulty measurement was another of Reeves' (1993, 1995) indicators of pseudoscience but in evaluation, particularly as it relates to "real world" educational contexts such as found in distance or distributed education, the problem is often more insidious. The burden of showing reliability and validity for any test not in general use is always upon the researcher (Burton & Magliaro, 1988). Many studies related to distance learning use "teacher-made achievement tests" which may, or may not have

reliabilities or validaties established. Perhaps worse than using a test which produces scores which are largely error or unrelated to the content however, is the fact that such tests are often used a part of a graded exercise. Graded exercises may cause people who tend to make A's and B's simply work harder to overcome any problem in the instruction. The potential lack of adequate tests are measurement problems. The potential difference in effort are violations of *ceteris paribus*.

In terms of statistics, many current researchers have argued that null hypothesis testing should be eliminated altogether (e.g. Carver, 1993) while others such as Thompson (1996) and Robinson & Levin (1997) would like to see such tests supplemented. Although there are differences, both camps tend to agree on two things: effect size and replication. Effect sizes should always be reported according to Thompson (1996) but, as Levin (1993) points out "to talk of effect sizes in the face of results that are not statistically significant does not make sense" (p. 379)

Replication refers to repeating essentially the same experiment multiple times. No finding should ever stand on a single study. It is worth noting however, that while some believe that such experiments can inform social science theory (e.g. Phillips, 1992), others (e.g. Salomon, 1991) believe that no matter how well constructed experimental and similar research approaches are, "they are based on a number of assumptions, none of which fit the study of whole classroom cultures" (p.13). We assume this would include distant classrooms and distributed cultures.

Finally we offer the following caveat related to accepting NSD studies as proof. Establishing a null is very much like the not guilty assumption of the US legal system. In both cases, the burden of proof is on overturning the assumption based on evidence. But failure to reject the null hypothesis means just that and nothing more; just as a legal finding of not guilty does not mean innocent. As Carver (1978) puts it:

What is the probability of obtaining a dead person given that the person was hanged? Obviously, it is very high, perhaps .97 or higher. Now, let us reverse the question. What is the probability that a person has been hanged, given that the person is dead? This time the probability will undoubtedly be very low, perhaps .01 or lower. No one would be likely to make the mistake of substituting the first estimate (.97) for the second (.01); that is to accept (.97) that a person has been hanged given that the person is dead. Even though this seems an unlikely mistake, it is exactly the kind of mistake that is made with interpretations of statistical significance testing (pp. 384 – 385).

Evaluation versus Research in Distance Education

While it may have been the intent of the investigators of comparison studies cited herein to create generalizable findings, the motivation behind the studies were most likely to obtain information about the success of local distance education programs. Appropriate uses of media comparisons for distance program evaluation are detailed as follows, as well as alternative methods and exemplary models.

Evaluation in distance education

Program evaluations in education frequently look to achievement as a measure of success, and sometimes through the use of comparison studies as an evaluation method. Smith & Glass (1987) call such inquiries comparative evaluations, as the studies assess the effectiveness of a product or program by pitting it against an alternative product or program that is designed to meet the same needs. However, such comparisons work best if the treatment group and control group are similar in identity and can be randomly assigned (Fitz-Gibbon & Morris, 1978), which is usually not the case in distance education. Participants in higher education distributed courses are typically non-traditional learners who cannot attend class at the originating institution, hence their enrollment in distance programs. Not only are these students different demographically, but they also possess other characteristics which vary from traditional college attendees, such as prior knowledge and experience and level of motivation (Verduin & Clark, 1991). In any case, if comparative evaluations can be designed to represent comparable groups of learners, the results of such studies must be published as local findings instead of generalizable contributions to the theoretical base of distance education.

Although student achievement is one common measure of distance program success, Keegan (1996), Holmberg (1989), and Thorpe (1988) recommend that program evaluators collect and report a number of other types of data to give the most exhaustive description of a distance education program. Saveyne & Box (1995) suggest the collection of information with regard to instructional design, participant attitudes (student and instructor), and implementation issues, such as technical quality, student support, etc. Keegan (1996) proposes a four-point evaluation scheme for distance programs which assesses 1) the quantity of learning achieved such as the number of new students served, attrition rates, time to program completion, etc.; 2) the quality of learning achieved measured by the effectiveness of the program in facilitating desired learning outcomes; 3) the status of the learning achieved indicated by the transferability of program coursework, recognition of degrees by employers or graduate institutions; and 4) the relative cost of the learning achieved acquired through the analysis of the cost-efficiency of distance programs relative to conventional programs, as well as the cost benefits of the distance program versus traditional programs (1996, pp.186-188). The case studies provided by Keegan (1996) are mindful examples of distance program evaluations, as they provide a thorough portrayal of program efforts through analysis of the aforementioned indicators. Another effective distance evaluation model can be found in the Flashlight Project (Ehrmann, 1994), an

BEST COPY AVAILABLE

effort by the Annenburg/Center for Public Broadcasting to help institutions of higher education assess their uses of instructional technology for distributed learning. If the intentions of investigators are to determine the effectiveness of distance education programs, these evaluation reports serve as exemplars due to their comprehensive approach.

Research in distance education

Those involved in the design, development, and implementation of distance education programs have access to a wealth of data from which to conduct valid research. For a summary of the existing literature base in distance education, see McIsaac & Gunawardena (1996) and Schlosser & Anderson (1994). Interestingly, both pieces highlight the need to move away from the continued use of media comparison studies toward more productive lines of inquiry. If researchers are driven to investigate the effects of delivery media, perhaps they will heed Reeves' (1995) advice and design instructional technology studies that will indeed improve education. Examples of research that have served to inform the development of effective distance learning experiences can be seen in Garrison (1990) and Gunawardena, Campbell Gibson, Dean, Dillon, & Hessmiller (1994). Garrison (1990) analyzed the ability of audioconferencing to provide necessary levels of interaction for feedback as well as for student satisfaction. Distance delivery media also afford varied levels of social presence, as found by Gunawardena, et al. (1994). Knowing how media convey information and allow individuals to interact are important considerations in the design of distance programming. Indeed, more researchers should leverage their involvement in distance education experiences to contribute to the knowledge base of the field. McIsaac and Gunawardena (1996) indicate that what is needed is "rich qualitative information or programmatic experimental research that would lead to the testing of research hypotheses" (p. 421). While determining the efficacy of distance programs is important to all stakeholders, investigators must ensure that such inquiries begin with valid questions and that the intentions behind the study are well-defined. Concurrently, it is equally important that editors of professional journals also distinguish between research and evaluation in distance education and communicate that distinction to the authors of their manuscripts.

Reference:

Barry, M., & Runyan, G. (1995). A review of distance-learning studies in the military. American Journal of Distance Education, 9(3), 37-47.

Burton, J. K., & Magliaro, S. G. (1988). Computer programming and generalized problem solving skills: In search of direction. In W. M. Reed & J. K. Burton (Eds.), Educational Computing and Problem Solving (pp. 63-90). New York: Haworth Press, Inc.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Carver, R. P. (1993). The case against statistical significance testing revisited. Journal of Experimental Education, 61(4), 287-292.

Clark, R. E. (1983). Reconsidering research on learning from media. Review of Educational Research, 53(4), 445-459.

Clark, R. E. (1994a). Media and method. Educational Technology Research and Development, 42(3), 7-10.

Clark, R. E. (1994b). Media will never influence learning. Educational Technology Research and Development, 42(2), 21-29.

Douglas, G. (1996). MLIS distance education at the University of South Carolina: Report of a case study. Journal of the American Society for Information Science, 47(11), 875-879.

Ehrmann, S. (1994). Project "Flashlight" planning grant: Final report. (HE030726). Washington, D.C.: Corporation for Public Broadcasting.

Fitz-Gibbon, C. T., & Morris, L. L. (1978). How to design a program evaluation. Beverly Hills, CA: Sage Publications.

Fox, M. F. (1996). Teaching a large enrollment, introductory geography course by television. Journal of Geography in Higher Education, 20(3), 355-365.

Garrison, D. R. (1990). An analysis and evaluation of audio teleconferencing to facilitate education at a distance. American Journal of Distance Education, 4(3), 13-24.

Gunawardena, C., Campbell Gibson, C., Dean, Dillon, C., & Hessmiller. (1994). Multiple perspectives on implementing inter-university computer conferencing. Paper presented at the Distance Learning Research Conference, San Antonio. TX.

Holmberg, B. (1989). Theory and Practice of Distance Education. London: Routledge.

Jonassen, D. H., Campbell, J. P., & Davidson, M. E. (1994). Learning with media: Restructuring the debate. Educational Technology Research and Development, 42(3), 31-39.

Kanner, J. H., Runyon, R. P., & Desiderato, O. (1958). Television in Army training. Audio-visual Communication Review, 6, 255-291.

Keegan, D. (1996). Foundations of Distance Education. (Third ed.). London: Routledge.

Keppel, G., & Zedeck, S. (1989). Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches. New York: W. H. Freeman.

Kozma, R. (1991). Learning with media. Review of Educational Research, 61(2), 179-211.

Kozma, R. B. (1994a). A reply: Media and methods. Educational Technology Research and Development, 42(3), 11-13.

Kozma, R. B. (1994b). Will media influence learning? Reframing the debate. Educational Technology Research and Development, 42(2), 7-19.

Levin, J. R. (1993). Statistical significance testing from three perspectives. Journal of Experimental Education, 61(4), 378-382.

McClusky, H. D., & McClusky, H. Y. (1924). Comparison of six modes of presentation of the subject matter contained in a film on the iron and steel industry and one on lumbering in the north woods. In F. N. Freeman (Ed.), Visual Education (pp. 229-274). Chicago: The University of Chicago Press.

McIsaac, M., & Gunawardena, C. (1996). Distance education. In D. H. Jonassen (Ed.), Handbook of Research for Educational Communications and Technology (pp. 403-437). New York: Simon & Schuster Macmillan.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science, 34, 103-115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and this slow progress of psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Monson, M. (1978). Bridging the Distance: An instructional guide to teleconferencing. Madison, WI: University of Wisconsin-Extension.

Moore, D. M., Myers, R. J., & Burton, J. K. (1994). Theories of multimedia and learning. In A. W. Ward (Ed.), Multimedia and Learning: A School Leader's Guide (pp. 29-41). Alexandria, VA: National School Boards Association.

Morrison, G. R. (1994). The media effects question: "Unresolvable" or asking the right question. Educational Technology Research and Development, 42(2), 41-44.

Musial, G., & Kampmueller, W. (1996). Two-way video distance education: Ten misconceptions about teaching and learning via interactive television. Action in Teacher Education, 17(4), 28-36.

Newlands, D., & McLean, A. (1996). The potential of live teacher supported distance learning: A case study of the use of audio conferencing at the University of Aberdeen. Studies in Higher Education, 21(3), 285-297.

Orey, M. A., Garrison, J. W., & Burton, J. K. (1989). A philosophical critique of null hypothesis testing. Journal of Research and Development in Education, 22(3), 12 - 21.

Phillips, D. C. (1992). The social scientist's bestiary: A guide to fabled threats to, and defenses of, naturalistic social science. Oxford: Pergamon Press.

Popper, K. R. (1968). Conjectures and refutations: The growth of scientific knowledge. New York: Harper Torchbooks.

BEST COPY AVAILABLE

Reeves, T. C. (1993). Pseudoscience in computer-based education: The case of learner control. Journal of Computer-Based Instruction, 20(2), 39-46.

Reeves, T. C. (1995). Questioning the questions of instructional technology. Paper presented at the National Convention of the Association for Educational Communications and Technology, Anaheim, CA.

Reiser, R. A. (1994). Clark's invitation to the dance: An instructional designer's response. Educational Technology Research and Development, 42(2), 45-48.

Reiss, J., Cameon, R., Matthews, D., & Shenkman, E. (1996). Enhancing the role public health nurses play in serving children with special health needs: An interactive videoconference on public law 99-457 Part H. Public Health Nursing, 13(5), 345-352.

Robinson, D. H., & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. Educational Researcher, 26(5), 21-26.

Ross, S., & Morrison, G. (1996). Experimental research methods. In D. H. Jonassen (Ed.), Handbook of Research for Educational Communications and Technology (pp. 1148-1170). New York: Simon & Schuster Macmillan.

Russell, T. (1997). The "No Significant Difference" phenomenon as reported in 248 research reports, summaries & papers . URL: http://www2.ncsu.edu/oit/nsdsplit.htm: North Carolina State University.

Saettler, P. (1968). History of Instructional Technology. New York: McGraw-Hill.

Salomon, G. (1991). Transcending the qualitative-quantitative debate: the analytic and systemic approaches to educational research. Educational Researcher, 20(6), 10-18.

Saveyne, W., & Box, C. (1995). Evaluation techniques. In B. Hakes, J. Cochenhour, L. Rezabek, & G. Sachs (Eds.), Compressed Video for Instruction: Operations and Applications (pp. 141-160). Washington, DC: Association for Educational Communications and Technology.

Schlosser, C., & Anderson, M. (1994). Distance education: Review of the literature (ISBN-0-89240-071-4). Ames, IA: Iowa State University.

Shrock, S. A. (1994). The media influence debate: Read the fine print, but don't lose sight of the big picture. Educational Technology Research and Development, 42(2), 49-53.

Smith, M. L., & Glass, G. V. (1987). Research and Evaluation in Education and the Social Sciences. Englewood Cliffs, New Jersey: Prentice-Hall.

Tennyson, R. D. (1994). The big wrench vs. integrated approaches: The great media debate. Educational Technology Research and Development, 42(3), 15-28.

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. Journal of Experimental Education, 61(4), 361-377.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Thorpe, M. (1988). Evaluating Open and Distance Learning. Essex, UK: Longman Group UK Limited.

Thyer, B. A., Polk, G., & Gaudin, J. G. (1997). Distance learning in social work education: A preliminary evaluation. Journal of Social Work Education, 33(2), 363-367.

Verduin, J. R., & Clark, T. A. (1991). Distance education: The foundations of effective practice. San Francisco, CA: Jossey-Bass.

Whetzel, D., Felker, D., & Williams, K. (1996). A real world comparison of the effectiveness of satellite training and classroom training. Educational Technology Research and Development, 44(3), 5-18.

Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA, *F*, *W*, and *F* statistics. Communications and Statistics, 15, 933-943.